AD-763 908

EFFECT OF PHOTO DEGRADATION ON INTERPRETER PERFORMANCE

Thomas E. Jeffrey

Army Research Institute for the Behavioral and Social Sciences
Arlington, Virginia

June 1973

Technical Research Note 245

AD

EFFECT OF PHOTO DEGRADATION ON INTERPRETER PERFORMANCE

Thomas E. Jeffrey
Army Research Institute for the Behavioral and Social Sciences

Abraham H. Birnbaum, Supervisory Project Director

U. S. Army

Research Institute for the Behavioral and Social Sciences

June 1973

48

U. S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

An activity of the Chief, Research and Development

J. E. UHLANER
Technical Director

R. A. ROOTH
COL, GS
Commander

NOTICES

DISTRIBUTION: Primary distribution of this report has been made by ARI. Please address correspondence concerning distribution of reports to: U. S. Army Research Institute for the Behavioral and Social Sciences, ATTN: RDMR-P, 1300 Wilson Boulevard, Arlington, Virginia 22209.

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U. S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) U.S. Army Research Institute for the Behavioral and Social Sciences, Arlington, VA	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

EFFECT OF PHOTO DEGRADATION ON INTERPRETER PERFORMANCE

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

5. AUTHOR(S) (First name, middle initial, last name)

Thomas E. Jeffrey

6. REPORT DATE June 1973	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS

8a. CONTRACT OR GRANT NO. b. PROJECT NO. DA R&D Proj. No. 20662704A721 c. Image Interpretation Displays d. a-13	9a. ORIGINATOR'S REPORT NUMBER(S) Technical Research Note 245 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Office, Chief of Research and Development, Washington, D. C.

13. ABSTRACT

Scale, haze, and image motion are judged to be the key factors in determining the interpretability of photographs encountered operationally, and images varying in these factors were produced by manipulating operational imagery in the laboratory. Image interpreters were given tasks of detecting and identifying military targets on these images. Their scores were used to assess interpretability as a function of scale, haze, and image motion. (U)

Results of the experiment showed: 1) When considered singly, scale, haze, and image motion had little effect on interpreter target detection performance. When two or more of these sources of degradation were present simultaneously, target detection performance deteriorated markedly. 2) In general, the decrease in target detection accuracy obtained with increased photo degradation appears to be due more to erroneous classification of non-targets as targets than to the misclassification of targets as, non-targets. 3) Target identification accuracy and completeness were significantly reduced by either unidimensional or multidimensional degrading conditions of the type used in this experiment. 4) When photographic scale was small, the effect of other sources of degradation on interpreter performance was significantly greater than when photographic scale was large. (U)

The findings provide direction for a continuing search for improved techniques for predicting the utility of aerial reconnaissance photographic missions and for guiding the establishment of mission requirements. (U)

DD FORM 1 NOV 65 1473 REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

Unclassified
Security Classification

Unclassified
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Aerial Surveillance						
*Aerial photographic imagery						
*Image interpreter performance						
*Photographic scale, haze, image motion						
*Photographic quality						
Ground-resolved distance (GRD)						
*Target detection						
*Target identification						
Detection accuracy						
Identification accuracy						
Identification completeness						

Unclassified
Security Classification

Technical Research Note 245

AD

EFFECT OF PHOTO DEGRADATION ON INTERPRETER PERFORMANCE

Thomas E. Jeffrey

Abraham H. Birnbaum, Supervisory Project Director

SYSTEMS INTEGRATION & COMMAND/CONTROL TECHNICAL AREA
Cecil D. Johnson, Chief

Submitted By:
Joseph Zeidner, Director
ORGANIZATIONS & SYSTEMS
RESEARCH LABORATORY

Approved By:
J. E. Uhlaner
TECHNICAL DIRECTOR

U. S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

Office, Chief of Research and Development
Department of the Army
1300 Wilson Boulevard, Arlington, Virginia 22209

June 1973

Army Project Number
2Q662704A721

Image Interpretation Displays a-13

ARI Technical Research Reports and Technical Research Notes are intended for sponsors of R&D tasks and other research and military agencies. Any findings ready for implementation at the time of publication are presented in the latter part of the Brief. Upon completion of a major phase of the task, formal recommendations for official action normally are conveyed to appropriate military agencies by briefing or Disposition Form.

FOREWORD

The SURVEILLANCE SYSTEMS Work Unit within the Army Research Institute (ARI) has as its objective the production of scientific data bearing on the extraction of information from surveillance displays and the efficient storage, retrieval, and transmission of this information within an advanced computerized image interpretation facility. Research results are used in future systems design and in the development of enhanced techniques for all phases of the interpretation process.

ARI research in this area is conducted as an in-house research effort augmented by contracts with organizations selected as having unique capabilities and facilities for research in aerial surveillance. The entire research program is responsive to requirements of Army RDT&E Project 2Q662704A721, Surveillance Systems, FY 1973 Work Program.

The ARI Work Unit, Surveillance Systems, is conducting research to determine how interpreter performance is affected by variations in the character of the image. A primary objective is to develop an instrument for use in evaluating imagery for interpretability--an image quality catalog, in effect. An analysis based on an analogy with signal detection concepts has been reported in Technical Research Report 1178, "Development of a psychophysical photo quality measure."

The research reported here was accomplished jointly by personnel of the Stanford Research Institute and by the Systems Integration and Command/Control Technical Area, Organization and Systems Research Laboratory of the U. S. Army Research Institute for the Behavioral and Social Sciences. The Institute, established 1 October 1972, as replacement for the U. S. Army Manpower Resources Research and Development Center, unifies in one enlarged organization all OCRD activities in the behavioral and social sciences area, including those conducted by the former Behavior and Systems Research Laboratory (BESRL) and the Motivation and Training Laboratory (MTL). The present publication reports on a special analysis of the data collected as a basis for development of the psychophysical photo quality measure and identifies atmospheric haze as an additional dimension to be included in such a measure.

J. E. UHLANER
Technical Director

EFFECT OF PHOTO DEGRADATIONS ON INTERPRETER PERFORMANCE

BRIEF

Requirement:

To identify photo dimensions frequently responsible for quality degradation of operationally obtained aerial reconnaissance photographic film and to assess their effect on the accuracy and completeness with which trained image interpreters can detect and identify tactical targets.

Procedure:

Factors contributing to poor quality photo-mission coverage were isolated by detailed examination of reconnaissance photography in several military film repositories. Three dimensions--photo scale (four levels), haze (three levels), and blur (four levels)--were selected for experimental manipulation. Of the 48 experimental conditions possible, 13 were selected for the research. Each of 13 aerial scenes was treated photographically to produce the 13 treatment conditions. The interpreter's task was to view a serially numbered set of circled areas on each photographic transparency and to judge whether targets were or were not contained in the circled area and to identify all targets. Scores on accuracy and completeness of target detection and identification were computed for each experimental subject. Means and standard deviations were obtained for every treatment condition. Analysis of variance was used to determine the statistical significance of treatment effects. Duncan's multiple range test was used to evaluate the statistical significance among the various means.

Findings:

When variations in photo scale, haze, and blur were present separately in photographic transparencies, there was little change in target detection performance. When two or more of these sources of degradation were present simultaneously, target detection deteriorated markedly.

Target identification accuracy and completeness were significantly reduced by either unidimensional or multidimensional degrading conditions of the type included in the investigation.

When photo scale was small, the effect of other sources of degradation on interpreter performance was significantly greater than when photo scale was large.

Degradation of overall target detection accuracy was due more to erroneous classification of non-targets as targets than to classification of targets as non-targets.

Utilization of Findings:

The findings of this technology base research provide direction for a continuing search for improved techniques for predicting the utility of aerial reconnaissance photographic missions and for guiding the G2 Air officer in establishing mission requirements.

In addition, findings will be useful in revision of BESRL's photo quality catalog from which measures of the interpretability of specific imagery are derived. Estimates are now based on comparison with catalog images varying in scale, sharpness, and scene complexity. Results here indicate that to these should be added variations in atmospheric haze as another index to the amount of information to be expected from interpretation of the imagery.

CONTENTS

TABLES

FIGURES

EFFECT OF PHOTO DEGRADATION ON INTERPRETER PERFORMANCE

BACKGROUND

The quality of the aerial photographs from which the image interpreter must extract intelligence information contributes importantly to the accuracy and completeness of his target detections and identifications. Although personal characteristics of the interpreter are important in determining the absolute level of performance in a given circumstance, losses due to individual differences are variable while those due to photo degradations tend to be more generalized, resulting in some performance loss for all interpreters. To predict the mean target detection and identification accuracy associated with photographs degraded by specified amounts, BESRL developed a photo quality catalog.[1] This catalog contains 231 photo transparencies which vary in scale, image sharpness, and scene complexity. Scene complexity refers to the amount of confusion introduced in the interpretation task by scene background. A low complexity scene would be characterized by flat open terrain where the target objects are readily distinguishable from the background. A scene of high complexity would contain many natural features such as rocks and vegetation that would make it exceedingly difficult to separate target from background. In estimating the interpretability of a photograph, the interpreter compares the photograph with the catalog transparencies. He finds the catalog image that he judges to match the photograph most closely in scale, sharpness, and scene complexity. The number of the catalog image is then used to enter the table provided with the catalog to obtain the predicted level of accuracy in target detection and target identification.

In subsequent research, BESRL sought to identify other photo dimensions that should be incorporated in the photo quality catalog. A search of operational reconnaissance film repositories led the investigators to conclude that photo scale, haze due to atmospheric attenuation, and blur resulting from camera movement or faulty image motion compensation were the most common causes of photographic degradation. Excellent quality photographic materials containing tactical targets were selected and degraded by darkroom procedures to produce photographs that represented the range of the experimental variables selected for the research. Four levels of photo scale, three levels of haze, and four levels of blur were specified for the experiment. Only 13 of the possible 48 treatment combinations were selected for the experiment because of the expense involved in treating imagery and testing large numbers of experimental subjects. Data were collected from trained image interpreters assigned to the 15th Military Intelligence Battalion, Aerial Reconnaissance and Support, located at Fort Bragg, North Carolina. These data

---

[1] Brainard, R. W., L. J. Lopez, G. N. Ornstein, and R. Sadacca. Development and evaluation of a catalog technique for measuring image quality. Behavior and Systems Research Laboratory Technical Research Report 1150. Arlington, Virginia. August, 1966.

were reduced and analyzed using techniques and procedures developed in signal detection theory. A simple model to predict target detection performance from the ground resolved distance (GRD) estimate for the photograph was developed.[2] The degree of agreement between predicted performance and empirical observation was examined using rank-difference correlation.

PURPOSE

The present purpose was to re-analyze data from the development of the psychophysical measure using different statistical procedures. The present treatment yields results which can be stated in terms consistent with those of other reports of research conducted in ARI's surveillance research program. A reader familiar with the research literature of aerial surveillance but unacquainted with the terms used in signal detection theory--receiver-operating-characteristics or ROC curves, for example--may find the present conceptualization more in keeping with specific interest in image interpretation than that presented in the earlier report on this research. Specific objectives of the present analysis were:

1. To determine the mean detection accuracy for the 13 treatment conditions. (Detection accuracy and completeness are equivalent indexes when the subjects are required to respond to a fixed set of annotated locations on the imagery.)

2. To determine mean target identification accuracy for each of the 13 treatment conditions.

3. To determine the mean target identification completeness for the 13 treatment conditions.

4. To determine separately for target and non-target annotations the mean detection accuracy/completeness.

METHOD

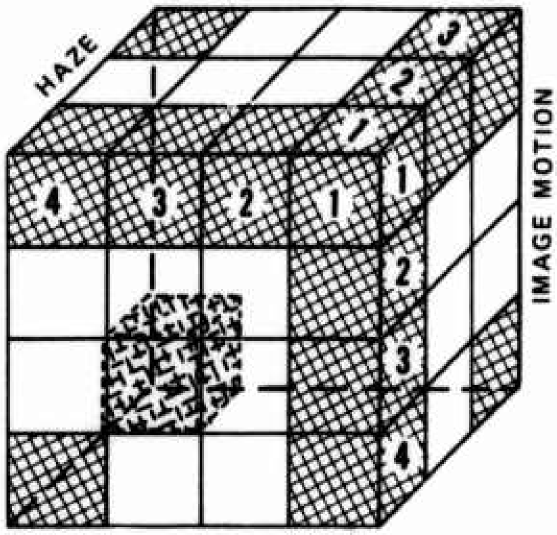
Experimental Design

Four levels of photo scale, three levels of haze, and four levels of image motion were established for the three independent measures of the experiment. Figure 1 shows a schematic representation of the research design. The total number of treatment conditions possible are 48 but because of the amount of work and expense involved in preparing imagery for all possible conditions only 13 were selected for the actual experiment. The

---

[2] Clarke, F. R., R. L. Welch, and T. E. Jeffrey. Development of a psychophysical photo quality measure. Army Research Institute Technical Research Report 1178. Arlington, Virginia. ______________ 1972.

13 experimental conditions chosen are indicated in Figure 1 by the crosshatched cells and Table 1 gives the level of scale, haze, and image motion for each.



LEGEND

Level	Photo Scale Nominal Value
1	1:2,000
2	1:4,000
3	1:8,000
4	1:12,000

Effect of Haze: Contrast Ratio	
Level	Ratio
1	40:1 (none)
2	4:1 (light)
3	2:1 (heavy)

Image Motion at Focal Plane (mm)	
Level	Movement
1	None
2	0.025
3	0.050
4	0.100

Figure 1. Schematic of Experimental Design

Table 1

EXPERIMENTAL CONDITION CODES AND INDEPENDENT VARIABLE LEVELS

Experimental Condition		Approximate[a] Scale	Haze	Image Motion
Number	Code			
1	1-1-1	1:2,000	None	None
2	2-1-1	1:4,000	None	None
3	3-1-1	1:8,000	None	None
4	4-1-1	1:12,000	None	None
5	1-2-1	1:2,000	1:4 Contrast	None
6	1-3-1	1:2,000	1:2 Contrast	None
7	1-1-2	1:2,000	None	.025mm
8	1-1-3	1:2,000	None	.050mm
9	1-1-4	1:2,000	None	.100mm
10	1-3-4	1:2,000	1:2 Contrast	.100mm
11	4-1-4	1:12,000	None	.100mm
12	4-3-1	1:12,000	1:2 Contrast	None
13	4-3-4	1:12,000	1:2 Contrast	.100mm

[a]See Table A-1 in Appendix A for actual scale values.

Development of Experimental Imagery

Fourteen large-scale, good-quality, negative transparencies were selected as the basic photographic imagery from which the experimental stimuli were to be prepared. Each 9 x 9-inch transparency depicted a unique scene. Thirteen of these scenes were for the collection of response data; the fourteenth was used as a practice scene to acquaint the subjects with the experimental task.

Target and non-target objects or areas were circled (annotated) on each of the 14 original negatives. The per scene average was about 15 annotations--8 containing targets and 7 without targets. Since the

target annotations could contain multiple targets, the average number of targets per scene exceeded the mean number of target annotations. There was an average of about 15 targets per scene for the 13 experimental images.

By photographic techniques the photo scale, haze, and blur were varied separately and in combination to produce positive transparencies for each of the 13 treatment conditions for every scene. Photo scale was varied by standard photo reduction methods. The haze effect was obtained by fogging the film using a beam splitter. Image movement effect was produced by moving the film easel at a controlled rate during exposure of the film. The practice image was reproduced at 1:9,600 scale and was without haze or blur. Ten annotations were present on this practice image.

The complete set of imagery consisted of 169 unique images--13 scenes with each at 13 treatment conditions. Multiple copies of the practice image were produced. Three complete sets of the experimental imagery were prepared so that separate packages of stimulus materials could be made up in which scene and treatment conditions were varied. Each Envelope contained one image of each of the 13 scenes, each scene produced under a unique treatment condition. The practice image was contained in a small envelope, and each of the stimulus images was numbered to permit ready identification.

Sample

Image interpreters assigned to the 15th Military Intelligence Battalion, Aerial Reconnaissance and Support located at Fort Bragg, North Carolina served as the experimental subjects. The men participating were mostly recent graduates of the Image Interpretation Course conducted at the U. S. Army Intelligence School then located at Fort Holabird, Maryland. Records from 26 of the 48 men tested were used in the present analysis.

Data Collection

Men were tested in groups of 13. Each man was provided a light table, 7-power tube magnifier, pencils, response booklet, and an envelope containing the experimental imagery. A target list like that appearing in Table 2 completed the number of items furnished.

The experimenter instructed the group to fill out the biographical data requested on the cover sheet of the response booklet, and, after all had completed this step, asked them to take out the practice image and place it on the light table. In a step-by-step sequence, the subjects were instructed in the procedure they were to follow in examining each annotation and in writing their responses in the answer booklet. After completion of the practice image and the resolution of all questions posed by the interpreters concerning the task, the experimenter proceeded with the administration of the experimental task. Rest breaks

Table 2

TARGET LIST

Targets	Nontargets
Vehicles (Utility, Commo, etc.)	Bushes, trees, logs, etc.
Truck, 2-1/2 ton	Old building foundations
Truck, 3/4 ton	
Truck, 1/4 ton	Aircraft shadows
Semitrailer, tank, gasoline	
Tow truck	Vehicle tracks
Tractor and semitrailer (van)	
	Crates, boxes
Armor	
Tank	Farm buildings
APC	
	Farm vehicles
Trailer	
1-1/2 ton and tank	Civilian vehicles on highways
3/4 ton	
1/4 ton	Livestock
Guns	
Howitzer (self-propelled)	
Howitzer (towed)	
Tents	
Large, CP	
Medium, CP, squad	
Small, pup	
Canvas, shelter, ammo	
Latrines	
Shower points	
Foxholes, one and two man	
Weapons pits	
Helicopters, utility	
Personnel	
Semipermanent and permanent buildings of military design, such as quonset, butler, etc.	

were given between the administration of successive images. Response booklets and experimental imagery were collected after completion of the final image. Four separate groups of 13 subjects each completed the task. Table A-2 shows the order of treatment conditions for the 26 men whose responses were analyzed for the present purpose.

Dependent Variables

The dependent variables included measures of target detection and target identification. The correctness with which the subject detected the presence or absence of targets in the annotations was used to derive indexes of detection performance. Figure 2 is a schematic presentation of the categories into which the responses of the subject were classified.

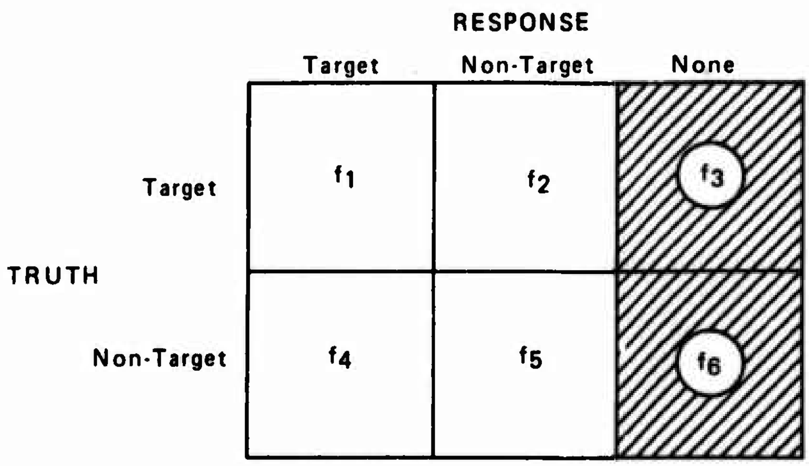


Figure 2. Response categories

The ratio measures, one for accuracy and one for completeness, were derived as measures of detection performance:

Detection accuracy. Number of annotations correctly classified, expressed as a ratio of all responses made by the subject:

$$\text{Detection accuracy} = \frac{(f_1 + f_5)}{(f_1 + f_5) + (f_2 + f_4)}$$

Detection completeness. Number of annotations correctly classified, expressed as a ratio of the total number of annotations in the imagery.

$$\text{Detection completeness} = \frac{(f_1 + f_5)}{(f_1 + f_5) + (f_2 + f_4) + (f_3 + f_6)}$$

All subjects were required to respond to all annotations; therefore, $f_3 = f_6 = 0$, as the value of the frequencies in the shaded cells of Figure 2. Equations (1) and (2) are thereby reduced to the same expression and the results for detection performance are reported by a single index.

Target identification performance was determined by the subject's ability to name the targets in annotations that he correctly classified as being target annotations. In Figure 2, these are in the cell labeled f_1. For target annotations properly classified as target annotations, the subject could identify the targets correctly (R), or he could make the following errors: misidentify the target (W_m), fail to identify a target and thus omit reporting an identification (0), or give a target identification for a non-target, thereby inventing a target (W_i). Any targets reported by the subject for annotations classified in the cell labeled f_4 had to be of the inventive type of wrong response. Targets actually present in target annotations erroneously classified by the subject as non-target annotations and falling in the cell labeled f_2 in Figure 2 were scored as omissions. The two indexes for target identification performance were:

<u>Target identification accuracy</u>. Number of correct identifications, expressed as a ratio of the total number of target identifications reported.

$$\text{Target identification accuracy} = \frac{R}{R + W_m + W_i}$$

<u>Target identification completeness</u>. Number of correct identifications, expressed as a ratio of the total number of targets present in the imagery.

$$\text{Target identification completeness} = \frac{R}{R + W_m + 0}$$

Statistical Computations

The basic data required to obtain the frequencies indicated in Figure 2 were obtained by scoring the response booklets of the 26 interpreter subjects. The correctness of their target identifications, the number of misidentifications, inventions, and omissions were determined. For each of the 26 subjects, the accuracy of detection and the accuracy and completeness of target identification were computed. Tables B-1, B-4, and B-7 list the values for these indexes of performance for each interpreter subject for each of the 13 treatment conditions. Tables B-2, B-5, and B-8 list the same indexes of performance for each interpreter for each of the unique image scenes. Means and standard deviations for the three indexes of performance were computed and are given

in Table 3. Analysis of variance was used to test the statistical significance of treatment effects, and Duncan's multiple range test[3] was used to test the difference between mean performance among the 13 treatment conditions.

RESULTS AND DISCUSSION

Detection Accuracy

For the best circumstance, the average interpreter correctly classified target and non-target annotations in about 80 percent of the cases. The detection accuracy column of Table 3 shows that for treatment condition (1-1-1)--describing imagery of the largest scale with no degradation due to atmospheric attenuation or image movement--detection accuracy was .80. The poorest detection accuracy occurred with treatment condition (4-3-1)--imagery of the smallest scale, maximum haze, but no image movement. Here, detection accuracy was .62 and indicates that the average interpreter classified the annotations correctly about 62 percent of the time.

A very natural question arises concerning the statistical significance of such differences. Does the mean performance of these 26 interpreters vary significantly as a result of the treatment conditions used? To answer this logical question, the variance of the treatment by subject score matrix appearing at Table B-1 was analyzed. There were no true replications across subjects for these data. In the experiment, each image interpreter was presented the image scenes in precisely the same order. This procedure facilitated the conduct of the experiment and avoided the possibility that one subject might obtain information about a subsequent scene from one of his fellow subjects. Even with the scene order fixed, the number of orders in which 13 treatment conditions can be presented is very large. Table A-2 shows the order in which the treatment conditions were presented to each of the 26 subjects for the practice image and the 13 test scenes.

The analysis of variance summary appears in Table B-3. Main effects--subjects, images, and experimental conditions--are significant beyond the .01 level. A test of the differences among all possible pairs of treatment means was made using Duncan's multiple range test (Table 4). The following generalizations appear warranted: Interpreter ability to distinguish target and non-target objects at specified locations on an image is not significantly reduced when the three degrading factors employed in this experiment are introduced singly in treating the imagery. However, with one exception, when these factors

---

[3] Edwards, A. L. Experimental design in psychological research. New York: Holt, Rhinehart, and Winston, 1963, 236 ff.

Table 3

MEAN DETECTION AND IDENTIFICATION ACCURACY AND COMPLETENESS

TREATMENT COMBINATION	DETECTION ACCURACY= COMPLETENESS		IDENTIFICATION			
			ACCURACY		COMPLETENESS	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
1-1-1	.796	.152	.564	.254	.635	.256
2-1-1	.770	.161	.467	.230	.521	.198
3-1-1	.743	.141	.392	.183	.488	.204
4-1-1	.736	.131	.374	.261	.349	.209
1-2-1	.752	.140	.422	.266	.439	.274
1-3-1	.774	.177	.420	.256	.370	.227
1-1-2	.771	.183	.445	.239	.503	.283
1-1-3	.749	.172	.412	.271	.433	.261
1-1-4	.764	.152	.402	.233	.448	.244
1-3-4	.738	.142	.369	.254	.358	.237
4-1-4	.670	.174	.242	.231	.222	.231
4-3-1	.622	.191	.189	.205	.162	.174
4-3-4	.632	.154	.196	.192	.150	.179

were used in combination to degrade the imagery, detection, performance deteriorated significantly. The one exception, treatment condition (1-3-4), appears to indicate that large image scale may offset the effects produced by the other two degrading factors. Imagery produced under this treatment condition was of the large scale, about 1:2,000, with maximum haze effect and greatest blurring due to image movement. Mean detection performance for imagery degraded in this fashion was not significantly poorer than that for imagery degraded in only one dimension or not degraded at all.

The three treatment conditions that produced the greatest loss in detection performance were all at the smallest scale, about 1:12,000. While no data were available for intermediate photo scales coupled with degradations produced by simulated atmospheric attentuation and blurring due to image movement, it seems reasonable to assume that when image scale is small, any additional loss in image quality brought about by other degrading factors will be accompanied by a significant reduction in detection accuracy. The 13 treatment conditions selected for the present experiment have provided some evidence concerning the effect of these factors on detection accuracy.

Table 4

DUNCAN'S MULTIPLE RANGE TEST FOR DIFFERENCES AMONG

TREATMENT MEANS FOR DETECTION ACCURACY PERFORMANCE

Treatment Code	Means	4-3-1	4-3-4	4-1-4	4-1-1	1-3-4	3-1-1	1-1-3	1-2-1	1-1-4	2-1-1	1-1-2	1-3-1	1-1-1	Shortest Significant Ranges		
	Means	.622	.632	.670	.736	.738	.743	.749	.752	.764	.770	.771	.774	.796	α = .05		α = .01
4-3-1	.622	---			$\underline{.114}$	$\underline{.116}$	$\underline{.121}$	$\underline{.127}$	$\underline{.130}$	$\underline{.142}$	$\underline{.148}$	$\underline{.149}$	$\underline{.152}$	$\underline{.174}$	R_2	.070	.092
4-3-4	.632		---		$\underline{.104}$	$\underline{.106}$	$\underline{.111}$	$\underline{.117}$	$\underline{.120}$	$\underline{.132}$	$\underline{.138}$	$\underline{.139}$	$\underline{.142}$	$\underline{.164}$	R_3	.074	.096
4-1-4	.670			---				.079	.082	.094	.100	.101	.104	$\underline{.126}$	R_4	.076	.098
4-1-1	.736				---										R_5	.078	.100
1-3-4	.738					---									R_6	.079	.102
3-1-1	.743						---								R_7	.080	.103
1-1-3	.749							---							R_8	.081	.104
1-2-1	.752								---						R_9	.082	.105
1-1-4	.764									---					R_{10}	.083	.106
2-1-1	.770										---				R_{11}	.084	.107
1-1-2	.771											---			R_{12}	.084	.107
1-3-1	.774												---		R_{13}	.085	.108
1-1-1	.796													---			

Significant mean differences are listed above the diagonal (Difference = Column Mean − Row Mean).

Statistical significance level of differences is indicated by an underscore where $P \leq .01$ and without an underscore where $P \leq .05$.

Undoubtedly, the foregoing is known to those who plan operational aerial surveillance missions. If point targets are to be detected, the altitude of the aircraft, focal length of the lens, time of day, amount of turbulence, and so forth are considered as the mission is planned. After the mission is flown, the suitability of the imagery acquired can be judged prior to interpretation. If the image scale is small and the imagery degraded by factors other than scale, the G2 Air Officer may decide to have the mission re-flown immediately in order to meet mission requirements.

Identification Accuracy

Table B-4 lists the identification accuracy scores of the 26 interpreters for each of the 13 treatment conditions and Table B-5 presents similar scores for the 13 image scenes. Table B-6 summarizes the analysis of variance of these data. Main effects--subjects, images, and experimental conditions--were statistically significant at better than the one percent level.

The differences in mean performance among the 13 treatment conditions were compared using Duncan's multiple range test (Table 5). Entries are for those treatment conditions where the differences between treatment means are statistically significant at $P \leq .05$ or better. Mean identification accuracy for the treatment condition yielding the best imagery (code 1-1-1)--largest scale, without haze, and without blur--was significantly greater than that obtained under all other treatment conditions. Any reduction in quality--single or multi-dimensional--significantly decreased identification accuracy of the interpreter.

Without exception, results for treatment conditions in which only one dimension was less than optimal followed the same pattern. Mean performance for these degrading conditions involving a single factor did not differ significantly from the mean performance obtained under other single-factor degrading conditions. However, the mean performance for these single-factor degradations differed significantly from the mean performance obtained when the smallest scale imagery was degraded on one or two additional dimensions. Finally, for large scale imagery such as that produced under treatment condition (code 1-3-4)--largest scale, maximum haze, and blur--mean performance was significantly better than when small scale imagery was degraded by haze or by haze and blur.

Identification Completeness

Table B-7 lists the identification completeness scores for the 26 subjects for each of the 13 treatment conditions and Table B-8 gives similar scores for these men for each of the 13 image scenes. Table B-9 summarizes the analysis of variance for these data. Main effects--

Table 5

DUNCAN'S MULTIPLE RANGE TEST FOR DIFFERENCES AMONG

TREATMENT MEANS OF IDENTIFICATION ACCURACY PERFORMANCE

Treat-ment Code	Means	4-3-1	4-3-4	4-1-4	1-3-4	4-1-1	3-1-1	1-1-4	1-1-3	1-3-1	1-2-1	1-1-2	2-1-1	1-1-1	Shortest Significant Ranges $\alpha = .05$		$\alpha = .01$
	Means	.189	.196	.242	.369	.374	.392	.402	.412	.420	.422	.445	.467	.564			
4-3-1	.189	---			.180	.185	.203	.213	.223	.231	.233	.256	.278	.375	R_2	.096	.126
4-3-4	.196		---		.173	.178	.196	.206	.216	.224	.226	.249	.271	.368	R_3	.101	.131
4-1-4	.242			---	.127	.132	.150	.160	.170	.178	.180	.203	.225	.322	R_4	.104	.135
1-3-4	.369				---									.195	R_5	.107	.137
4-1-1	.374					---								.190	R_6	.109	.139
3-1-1	.392						---							.172	R_7	.110	.141
1-1-4	.402							---						.162	R_8	.111	.143
1-1-3	.412								---					.152	R_9	.113	.144
1-3-1	.420									---				.144	R_{10}	.114	.145
1-2-1	.422										---			.142	R_{11}	.115	.146
1-1-2	.445											---		.119	R_{12}	.115	.147
2-1-1	.467												---	.097	R_{13}	.116	.148
1-1-1	.564													---			

Significant mean differences are listed above the diagonal (Difference = Column Mean - Row Mean).

Statistical significance level of differences is indicated by an underscore where $P \leq .01$ and without an underscore where $P \leq .05$.

Table 6

DUNCAN'S MULTIPLE RANGE TEST FOR DIFFERENCES AMONG TREATMENT MEANS OF IDENTIFICATION COMPLETENESS PERFORMANCE

Treatment Code	Means	4-3-4	4-3-1	4-1-4	4-1-1	1-3-4	1-3-1	1-1-3	1-2-1	1-1-4	3-1-1	1-1-2	2-1-1	1-1-1	Shortest Significant Ranges	α = .05	α = .01
	Means	.150	.162	.222	.349	.358	.370	.433	.439	.448	.488	.503	.521	.635			
4-3-4	.150	---			$\underline{.199}$	$\underline{.208}$	$\underline{.220}$	$\underline{.283}$	$\underline{.289}$	$\underline{.298}$	$\underline{.338}$	$\underline{.353}$	$\underline{.371}$	$\underline{.485}$	R_2	.113	.149
4-3-1	.162		---		$\underline{.187}$	$\underline{.196}$	$\underline{.208}$	$\underline{.271}$	$\underline{.277}$	$\underline{.286}$	$\underline{.326}$	$\underline{.341}$	$\underline{.359}$	$\underline{.473}$	R_3	.119	.155
4-1-4	.222			---	.127	.136	.148	$\underline{.211}$	$\underline{.217}$	$\underline{.226}$	$\underline{.266}$	$\underline{.281}$	$\underline{.299}$	$\underline{.413}$	R_4	.123	.160
4-1-1	.349				---						$\underline{.139}$	$\underline{.154}$	$\underline{.172}$	$\underline{.286}$	R_5	.126	.163
1-3-4	.358					---					.130	.145	.163	$\underline{.277}$	R_6	.129	.165
1-3-1	.370						---					.133	.151	$\underline{.265}$	R_7	.131	.167
1-1-3	.433							---						$\underline{.202}$	R_8	.132	.169
1-2-1	.439								---					$\underline{.196}$	R_9	.134	.171
1-1-4	.448									---				$\underline{.187}$	R_{10}	.135	.172
3-1-1	.488										---			.147	R_{11}	.136	.173
1-1-2	.503											---		.132	R_{12}	.137	.174
2-1-1	.521												---	.114	R_{13}	.138	.175
1-1-1	.635													---			

Significant mean differences are listed above the diagonal (Difference = Column Mean - Row Mean).

Statistical significance level of differences is indicated by an underscore where P ≤ .01 and without an underscore where P ≤ .05.

images, subjects, and experimental conditions--were statistically significant at better than the one percent level.

The differences in mean performance among the 13 treatment conditions were compared using Duncan's multiple range test. Table 6 shows that identification completeness performance followed a pattern quite similar to that obtained for identification accuracy. For the best quality imagery used in the experiment (code 1-1-1), identification completeness was significantly superior to that obtained from any of the twelve experimental variants of the best condition.

For any of the single-factor degrading conditions, identification completeness was better than that obtained for imagery of the smallest scale subjected to additional degradation in one or both of the other two degrading dimensions. The largest scale imagery degraded maximally by haze and blur (code 1-3-4) gave results similar to those of the single degrading conditions. It appears that when the largest scale imagery is degraded by haze and blur, the decrement in identification completeness is significantly less than when the smallest scale imagery is degraded by either haze or blur or by both.

The pattern of significant mean differences for identification completeness differs from that obtained for identification accuracy in the following eight instances: 1) Imagery of scale 1:4,000 with no other degradation yielded better completeness performance than that obtained from imagery of 1:12,000 scale without additional degradation. 2) Imagery of 1:8,000 scale but no other degradation resulted in better identification completeness scores than that obtained with imagery of scale 1:12,000 but no other degradation. 3) Imagery of 1:4,000 scale as the only degrading factor gave better completeness results than was obtained from imagery of 1:2,000 scale degraded by maximum haze and blur. 4) Imagery of 1:8,000 scale as the only degrading factor resulted in better identification completeness than that obtained from imagery of 1:2,000 scale with maximum haze and blur. 5) Imagery of 1:4,000 scale and no other degradation results in more complete identification than is obtained with imagery of 1:2,000 scale with maximum haze effect but no blur. 6) Imagery of 1:2,000 scale, without haze but with the least appreciable amount of blur produced more complete responses than were obtained from 1:12,000 scale images without added degradation. 7) This same type of large scale as described in (6) was superior to imagery of the largest scale but with maximal amounts of haze and blur. 8) The same large scale imagery as described in (6) resulted in more complete performance than that obtained with imagery of the largest scale, maximal haze, but without blur. The small loss in quality resulting from the introduction of the smallest discrete amount of blurring did not produce any marked decrease in identification completeness. This result may have been due to the fact that the blurring effect was one-dimensional and the amount of movement was relatively small. The one-dimensional nature of the blurring effect was the result of the method used to simulate this dimension. The film on which the image was being copied was moved at controlled rates. Blurring of the image took place along the line of this movement.

All three experimental dimensions used in varying image quality in this experiment are seen to have produced significant differences in identification completeness within the range employed in this research. The effect of any combination of these degrading factors was more pronounced when the image scale was very small.

Absolute Levels of Performance

Identification Accuracy. In the preceding paragraphs the relative aspects of identification accuracy and their dependence on the various treatment conditions were discussed. One point of interest that should be discussed is the absolute level of identification accuracy attained in the experiment. Under the best condition of image quality, identification accuracy was no better than .56 for the average interpreter. This level of performance is not atypical from the results obtained in other surveillance research experiments. However, an examination of the factors operating in this specific experiment may help to explain why the absolute level of identification accuracy was not larger.

Identification accuracy as an index of performance is based on the number of target responses made by the interpreter. This number of responses is the denominator of the fraction and includes the number of correct identifications plus the number of target misidentifications plus the number of non-targets erroneously identified as targets (inventive responses). The numerator of the fraction is the number of correct identifications. The number of correct responses is directly dependent on the level of detail to which the targets must be identified. For this experiment, the interpreters were required to identify the targets rather precisely. Trucks, for example, were to be identified by tonnage. A response of "truck" was not scored as correct for an imaged object that was a 2 1/2-ton truck. This requirement for precision in naming target objects reduced the number of correct identifications and increased the number of misidentifications. The numerator of the identification accuracy index was thereby reduced while the denominator was increased. This is one of the factors operating to reduce the size of the identification accuracy index.

A second factor associated with the absolute level of identification accuracy relates to the number of non-target annotations that were wrongly judged to contain targets. The number of such erroneous detections in this particular research may be unduly large as a result of the way non-target areas were annotated. Here, the non-target annotations in the imagery were selected deliberately to include terrain features and man-made objects of the types interpreters frequently confuse with tactical targets. Objects such as rocks, rectangular outlines, highlighted tree crowns, shadows and wet spots on the road were annotated. The nature of the non-target annotations used in this experiment may have increased the likelihood that an interpreter would name the non-targets as target objects. The number of such inventive responses increased the denominator of the identification accuracy index and thereby makes the index smaller. These two factors may have been responsible for the absolute level of identification accuracy obtained in the present experiment.

Identification Completeness. The absolute level of identification completeness obtained by the subjects of this experiment merits comment. The index of identification for the best imagery (code 1-1-1) used in the experiment was about 64 percent. Why were fewer than two-thirds of the targets properly identified under the best of conditions? Insufficient working time might be advanced as a reasonable explanation. However, time was not responsible, since the interpreters were allowed enough time to complete each annotation. A second possible cause is the level of detail required for a correct response. In order to be scored as a correct identification, the interpreter had to identify the target rather explicitly. This is the same argument as that presented in the discussion for identification accuracy. This requirement for an exact name for the target reduces the number of correct responses and consequently the value of the numerator of the completeness index. A third possible cause deals with the nature of the stimulus material. Some of the annotations encircled multiple targets. If an annotation contained three M-48 tanks, the interpreter had to report all three in order to obtain full identification credit. The interpreters had been instructed to report all targets contained in each annotated area and had been informed that some annotations would contain multiple targets. They were not told which of the annotations actually contained the multiple targets. Previous research has shown that interpreters sometimes report one target of a cluster but fail to notice or fail to report the adjacent targets in the cluster. Therefore, the presence of multiple targets may have lowered the level of completeness attained by the subjects in this experiment. A fourth factor that may have been operating concerns the limit imposed by the level of detection completeness achieved by each interpreter. For each annotation, the interpreter judged whether a target was or was not present. For those annotations judged not to have a target, the interpreter made no identification. Therefore, the more actual target annotations the interpreter erroneously classified as non-target annotations, the lower his maximal identification completeness ceiling became. Maximal identification completeness would have been obtained had the interpreter classified every annotation as a target annotation and then have reported his best estimate of the identity of the real or imagined targets present in these annotations. To indicate how failure to classify annotations properly limits identification completeness, imagine that an interpreter judged that only 20 of the 104 target annotations contained in the imagery were target annotations and then correctly identified one target in each of the 20 annotations. His identification completeness score would be 20/194 (there were 194 targets present in the 104 target annotations) or about 10 percent. The extent to which proper classification of annotations limited identification completeness might be sought by referring to the values listed in Table 3. However, these figures refer to detection performance for target annotations and non-target annotations combined. The interpreter's performance in correctly classifying non-target areas and target areas were summed and expressed as a ratio of 200--the total number of target plus non-target annotations.

Detection Performance. Table 7 gives detection completeness for the target and non-target annotations separately and repeats the data from

Table 7

MEAN DETECTION ACCURACY (COMPLETENESS) BY ANNOTATION TYPE

Treatment Condition	Target Annotations		Non-Target Annotations		All Annotations	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
1-1-1	.935	.110	.666	.281	.796	.152
2-1-1	.930	.110	.633	.260	.770	.161
3-1-1	.938	.079	.500	.287	.743	.141
4-1-1	.824	.261	.577	.272	.736	.131
1-2-1	.916	.119	.569	.274	.752	.140
1-3-1	.912	.122	.667	.218	.774	.177
1-1-2	.931	.102	.601	.270	.771	.183
1-1-3	.915	.123	.631	.244	.749	.172
1-1-4	.924	.134	.616	.287	.764	.152
1-3-4	.859	.209	.582	.268	.738	.142
4-1-4	.658	.373	.556	.304	.670	.174
4-3-1	.691	.304	.452	.299	.622	.191
4-3-4	.699	.268	.499	.315	.632	.154
MEANS	.856	.224	.581	.284	.732	.169

Table 3 for the combined results. One striking feature can be seen in the performance for non-target annotations. As image quality was degraded, the correct classification of both target and non-target annotations declined. However, for non-target annotations, detection accuracy (completeness) dropped to chance performance or below. If the interpreter were to toss a coin for each annotation--heads, it's a target; tails, it's a non-target--one would expect that for a large number of such annotations he would classify about 50 percent correctly. Even for the best quality imagery, the mean performance for correct classification of non-target annotations was only 67 percent; that is, one-third of the annotations were classified as target annotations when in fact no targets were present.

The mean detection accuracy or completeness for target annotations was about 86 percent with a high of 94 percent for the best quality imagery and a low of 66 percent for one of the poorer quality image variants. It appears that as image quality is degraded the average interpreter is less able to detect target cues and signatures and, as a consequence, classifies more of the target annotations as being non-target annotations. However, under no circumstance in this experiment

did detection performance for target annotations deteriorate to chance level (50 percent) in this experiment. The analysis of variance summary for detection accuracy performance for target annotations appears in Table B-10 and that for non-target annotations in Table B-11. Main effects--subjects, images, and experimental conditions--are all significant beyond the .01 level. Differences among treatment condition means were tested using Duncan's multiple range technique and are reported in Table B-12 for target annotations and in Table B-13 for non-target annotations. The pattern of significant differences for target annotations is quite similar to that obtained for all annotations which appeared in Table 4. Results for non-target annotations seem to show that the interpreters found the classification task to be much more difficult for non-target objects. Detection performance dropped to chance level when only a moderate degree of photo degradation was introduced.

From the foregoing discussion, it seems that the level of detection accuracy (completeness) was not one of the limiting factors responsible for the modest level of target identification completeness obtained in the experiment. For example, the best quality imagery (code 1-1-1) the mean detection completeness for target annotations was 94 percent. Therefore, about 94 percent of the targets in the imagery were available to the interpreters for identification, but observed identification completeness was only 64 percent. For this reason, it seems that the explanation for the absolute level of identification completeness obtained must be attributed to the presence of multiple targets in the annotations and the difficulty of the identification task--the level of detail required in order to receive credit for a correct target identification.

The following observations sum up this discussion of absolute levels of performance:

The absolute level of detection accuracy was to a large extent determined by the difficulty of the non-target annotations. For about one-fourth of the experimental conditions, detection performance for non-target annotations was at chance level (50 percent).

The absolute level of identification accuracy appeared to be dependent on the level of detail required for a correct identification response and by the number of inventive errors made by the interpreters.

The absolute level of identification completeness seemed to be governed by the presence of multiple targets in the target annotations and by the level of detail demanded of the interpreter in order to obtain credit for a correct response.

The levels of performance obtained in this experiment do not apply directly to the operational situation. The imagery for the experiment was annotated, and subjects were paced through the imagery annotation by annotation--a "directed search" condition. The absolute performance levels for detection and identification may be very different from those which might have been obtained in a "free search" situation in which

there are no annotations and the interpreter must search for targets and identify the objects.

CONCLUSIONS

With respect to the effects of the degradation sources included in the experiment, the following conclusions appear warranted:

Any degradation of photo quality, unidimensional or multidimensional, significantly reduces the accuracy and completeness of target identification.

Unidimensional degradation of photo quality does not significantly reduce the level of detection performance (accuracy/completeness), whereas multidimensional degradation is associated with significant deterioration of detection performance.

In general, detection and identification performance for imagery degraded on only a single dimension is significantly superior to that for imagery degraded on more than one dimension.

The effect of degradation of photo quality by haze and blur, or both, is more pronounced for small scale imagery than for large scale imagery.

Performance as measured by all three dependent measures differed significantly between interpreters. Within interpreters, performance differed significantly by scene content (complexity) and by kind and amount of photo quality degradation.

The research was conducted for the purpose of identifying additional dimensions of photo quality that should be included in the development of the next generation of the BESRL photo quality catalog. The following results appear applicable to that goal: Haze effect produced operationally by atmospheric attenuation should be represented. While this source of degradation does not have an enormous effect on interpretation performance in isolation, it does result in significant deterioration of performance when coupled with other sources of degradation. As a dimension of photo quality, the effect of haze should be defined and quantified and its effect in interaction with scale, sharpness, and scene complexity determined so that it can be adequately covered in the catalog imagery.

APPENDIXES

EXPERIMENTAL CONDITIONS, SCORE MATRIXES,
AND ANALYSIS OF VARIANCE SUMMARIES

Appendix continued | Page

Table A-1

SCALE OF IMAGERY PRODUCED FOR TESTS

Image Number	Scale for Various Experimental Conditions					
	1[a]	2	3	4	5-10	11-13
1	1:2100	1:3810	1:8200	1:12,600	1:2100	1:12,600
2	1:2100	1:3810	1:8200	1:12,600	1:2100	1:12,600
3	1:2000	1:3650	1:8000	1:12,000	1:2000	1:12,000
4	1:2400	1:4360	1:9600	1:14,500	1:2400	1:14,500
5	1:2400	1:4360	1:9600	1:14,500	1:2400	1:14,500
6	1:2400	1:4360	1:9600	1:14,500	1:2400	1:14,500
7	1:2000	1:3650	1:8000	1:12,000	1:2000	1:12,000
8	1:3300	1:6000	1:13,250	1:20,000	1:3300	1:20,000
9	1:3200	1:5800	1:12,800	1:19,400	1:3200	1:19,400
10	1:3200	1:5800	1:12,800	1:19,400	1:3200	1:19,400
11	1:2000	1:3650	1:8000	1:12,000	1:2000	1:12,000
12	1:2000	1:3650	1:8000	1:12,000	1:2000	1:12,000
13	1:2000	1:3650	1:8000	1:12,000	1:2000	1:12,000

[a]Original negative scale.

Table A-2

EXPERIMENTAL CONDITIONS FOR IMAGE X EXAMINEE COMBINATIONS[a]

Ind. Code No.	Image Number 1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	3	1	12	4	9	6	2	7	10	3	5	13	8	11
2	3	2	13	5	10	7	3	8	11	4	6	1	9	12
3	3	3	1	6	11	8	4	9	12	5	7	2	10	13
4	3	4	2	7	12	9	5	10	13	6	8	3	11	1
5	3	5	3	8	13	10	6	11	1	7	9	4	12	2
6	3	6	4	9	1	11	7	12	2	8	10	5	13	3
7	3	7	5	10	2	12	8	13	3	9	11	6	1	4
8	3	8	6	11	3	13	9	1	4	10	12	7	2	5
9	3	9	7	12	4	1	10	2	5	11	13	8	3	6
10	3	10	8	13	5	2	11	3	6	12	1	9	4	7
11	3	11	9	1	6	3	12	4	7	13	2	10	5	8
12	3	12	10	2	7	4	13	5	8	1	3	11	6	9
13	3	13	11	3	8	5	1	6	9	2	4	12	7	10
14	3	13	1	7	8	9	3	5	6	10	2	4	11	12
15	3	11	12	13	1	7	10	8	9	5	6	2	3	4
16	3	3	4	11	12	13	5	1	7	8	9	6	10	2
17	3	10	2	3	4	11	8	12	13	1	7	9	5	6
18	3	5	6	10	2	3	1	4	11	12	13	7	8	9
19	3	8	9	5	6	10	12	2	3	4	11	13	1	7
20	3	1	7	8	9	5	4	6	10	2	3	11	12	13
21	3	12	13	1	7	8	2	9	5	6	10	3	4	11
22	3	4	11	12	13	1	6	7	8	9	5	10	2	3
23	3	2	3	4	11	12	9	13	1	7	8	5	6	10
24	3	6	10	2	3	4	7	11	12	13	1	8	9	5
25	3	9	5	6	10	2	13	3	4	11	12	1	7	8
26	3	7	8	9	5	6	11	10	2	3	4	12	13	1

[a]Experimental conditions are keyed as shown in Table 1.

Table B-1

DETECTION ACCURACY SCORES: SUBJECT BY TREATMENT CONDITION

Ind. Code No.	Treatment Condition Code 1-1-1	2-1-1	3-1-1	4-1-1	1-2-1	1-3-1	1-1-2	1-1-3	1-1-4	1-3-4	4-1-4	4-3-1	4-3-4	SUM
1	.533	1.000	.923	.750	1.000	.875	.667	.938	.812	.928	.950	.529	.900	10.805
2	.700	.667	.706	.538	.812	.917	.938	.938	.722	.688	.357	.800	.470	9.253
3	1.000	.800	.667	.882	.846	.750	.875	.833	.812	.944	.688	.786	.950	10.833
4	.750	.706	.400	.600	.765	.231	.083	.688	.812	.625	.667	.312	.643	7.282
5	.714	1.000	.765	.700	.600	.824	.625	.462	.417	.812	.562	.833	.625	8.939
6	.500	.857	1.000	.882	.700	.667	.692	.941	.875	.500	.688	.562	.556	9.420
7	.833	.500	.714	.850	.824	.800	.706	.600	.462	.750	.167	.688	.500	8.394
8	.875	.778	.625	.643	.950	.941	.600	.600	.882	.769	.875	.250	.562	9.350
9	.875	.688	.667	.500	.571	.850	.500	.824	.533	.765	.462	.562	.167	7.964
10	.667	.938	.688	.667	.312	.571	.941	.950	.900	.533	.706	.538	.562	8.973
11	.625	.417	.875	.688	.722	.375	.950	.714	.882	.700	.733	.647	.692	9.020
12	.769	.875	.667	.812	.812	.889	.928	.688	.950	.882	.800	.667	.588	10.327
13	.941	.769	1.000	.833	.857	.875	.938	.889	.857	.950	.765	.900	.733	11.307
14	.941	.750	.824	.900	.812	.928	.750	.625	.938	.692	.667	.950	.533	10.310
15	.812	.800	.889	.950	.615	.750	.812	.875	1.000	.824	.800	.353	.562	10.042
16	.938	.950	.667	.470	.882	.900	.692	.714	.833	.722	.812	.438	.750	9.768
17	.462	.882	.625	.562	.722	.900	.824	.667	.600	.600	.812	.562	.500	8.718
18	.824	.375	.812	.688	.733	.941	.778	.600	.750	.812	.500	.923	.833	9.569
19	.889	.688	.571	.923	.750	.750	.733	.950	.824	.750	.750	.529	.700	9.807
20	.667	.538	.583	.765	.750	.750	.938	.941	.750	.643	.900	.889	.750	9.864
21	.688	.882	.600	.667	.571	.538	.812	.312	.562	.333	.750	.667	.647	8.029
22	.938	.889	.950	.800	.833	.882	.928	.812	.615	.800	.529	.688	.562	10.226
23	1.000	.800	.824	.875	.700	1.000	.750	.462	.941	.900	.688	.375	.750	10.065
24	.917	.875	.750	.750	.950	.733	.800	.941	.722	.706	.500	.714	.769	10.127
25	.900	.812	.750	.786	.764	.688	.900	.778	.667	.688	.538	.417	.529	9.217
26	.950	.786	.769	.667	.688	.812	.882	.733	.750	.875	.765	.600	.611	9.888
TOTALS	20.708	20.022	19.311	19.148	19.541	20.137	20.042	19.475	19.868	19.191	17.431	16.179	16.444	247.497

Table B-2

DETECTION ACCURACY SCORES: SUBJECT BY IMAGE

Ind. Code No.	Image Number 1	2	3	4	5	6	7	8	9	10	11	12	13	SUM
1	.533	.529	.750	.812	.875	1.000	.938	.928	.923	1.000	.900	.667	.950	10.805
2	.667	.470	.812	.688	.938	.706	.938	.357	.538	.917	.700	.722	.800	9.253
3	.667	1.000	.750	.688	.875	.882	.812	.786	.846	.833	.800	.944	.950	10.833
4	.600	.706	.688	.312	.812	.765	.625	.643	.231	.083	.400	.667	.750	7.282
5	.600	.765	.625	.625	.812	.824	.562	.714	.462	.417	.700	.833	1.000	8.939
6	.667	.882	.875	.500	.688	.941	.562	.857	.692	.500	.700	.556	1.000	9.420
7	.600	.824	.750	.500	.688	.706	.500	.714	.462	.167	.800	.833	.850	8.394
8	.600	.941	.875	.625	.562	.882	.875	.643	.769	.250	.600	.778	.950	9.350
9	.533	.824	.562	.500	.875	.765	.688	.571	.462	.167	.500	.667	.850	7.964
10	.533	.941	.562	.312	.938	.706	.688	.571	.538	.667	.900	.667	.950	8.973
11	.733	.882	.625	.375	.875	.647	.688	.714	.692	.417	.700	.722	.950	9.020
12	.667	.882	.875	.688	.812	.588	.812	.928	.769	.667	.800	.889	.950	10.327
13	.733	.765	1.000	.938	.857	.941	.875	.857	.769	.833	.900	.889	.950	11.307
14	.533	.941	.625	.750	.938	.824	.812	.928	.692	.750	.900	.667	.950	10.310
15	.800	.353	.562	.812	.875	.824	.812	1.000	.615	.750	.800	.889	.950	10.042
16	.667	.470	.812	.438	.750	.882	.938	.714	.692	.833	.900	.722	.950	9.768
17	.600	.882	.625	.562	.812	.824	.562	.500	.462	.667	.600	.722	.900	8.718
18	.733	.941	.812	.375	.812	.824	.688	.500	.923	.833	.600	.778	.750	9.569
19	.733	.824	.750	.750	.750	.529	.688	.571	.923	.750	.700	.889	.950	9.807
20	.667	.941	.938	.750	.750	.765	.750	.643	.538	.583	.900	.889	.750	9.864
21	.667	.647	.688	.312	.812	.882	.562	.571	.538	.333	.600	.667	.750	8.029
22	.800	.529	.688	.562	.938	.882	.812	.928	.615	.833	.800	.889	.950	10.226
23	.800	.824	.875	.688	.375	.941	.750	1.000	.462	.750	.700	1.000	.900	10.065
24	.733	.706	.875	.750	.750	.941	.500	.714	.769	.917	.800	.722	.950	10.127
25	.667	.764	.688	.688	.812	.529	.750	.786	.538	.417	.900	.778	.900	9.217
26	.733	.882	.750	.688	.812	.765	.875	.786	.769	.667	.600	.611	.950	9.888
TOTALS	17.266	20.115	19.437	15.688	20.793	20.765	19.062	18.924	16.689	16.001	19.200	20.057	23.500	247.497

Table B-3

ANALYSIS OF VARIANCE SUMMARY: DETECTION ACCURACY SCORES

Source of Variation	Sum of Squares	df	Mean Square	F	$F_{.95}$	$F_{.99}$
Between:						
Subjects	1.71793	25	.06872	4.16**	1.55	1.84
Within:						
Images	2.21050	12	.18421	11.15**	1.79	2.25
Conditions	.95004	12	.07917	4.79**	1.79	2.25
Residual	4.75798	288	.01652	----	----	----
TOTAL	9.63645	337	------	----	----	----

**Means significantly different, $P \leq .01$.

Table B-4

IDENTIFICATION ACCURACY SCORES: SUBJECT BY TREATMENT CONDITION

Ind. Code No.	Treatment Condition Code 1-1-1	2-1-1	3-1-1	4-1-1	1-2-1	1-3-1	1-1-2	1-1-3	1-1-4	1-3-4	4-1-4	4-3-1	4-3-4	SUM
1	.455	.737	.500	.417	1.000	.400	.231	.556	.333	1.000	.615	.200	.250	6.694
2	.636	.462	.400	.125	.588	.500	.667	.619	.556	.333	.000	.579	.083	5.548
3	.783	.500	.357	.200	.000	.200	.450	.000	.385	.556	.333	.000	.312	4.076
4	.600	.355	.364	.375	.385	.000	.077	.111	.280	.294	.600	.176	.000	3.617
5	.538	.684	.321	.444	.461	.500	.625	.000	.125	.364	.200	.536	.214	5.012
6	.444	.125	.647	.444	1.000	.545	.000	.304	.667	.000	.000	.100	.250	4.526
7	.640	.222	.182	.619	.100	.333	.500	.111	.091	.273	.000	.071	.083	3.225
8	.600	.727	.273	.000	.684	.350	.545	.500	.600	.000	.500	.000	.267	5.046
9	.567	.438	.278	.077	.357	.273	.231	.242	.080	.207	.071	.154	.000	2.975
10	.143	.571	.428	.667	.154	.444	.476	.842	.556	.353	.308	.000	.462	5.404
11	.600	.125	.500	.412	.304	.267	.667	.333	.174	.400	.545	.200	.000	4.527
12	.000	.385	.200	.364	.364	.700	.667	.454	.650	.182	.286	.400	.062	4.714
13	.778	.200	.727	.000	.500	.444	.625	.800	.857	.941	.143	.250	.438	6.703
14	.850	.250	.379	.800	.462	1.000	.454	.500	.478	.000	.727	.571	.600	7.071
15	.444	.833	.577	1.000	.400	.000	.636	.571	.143	.308	.583	.000	.000	5.495
16	.667	.800	.000	.222	.562	.714	.000	.750	.500	.652	.111	.182	.167	5.327
17	.000	.310	.391	.286	.543	.824	.475	.100	.500	.214	.048	.048	.000	3.739
18	.667	.200	.333	.235	.571	.103	.480	.500	.522	.500	.000	.000	.000	4.111
19	.783	.429	.444	.000	.200	.222	.214	.933	.462	.500	.000	.091	.500	4.778
20	.438	.111	.167	.462	.474	.308	.583	.318	.333	.454	.400	.682	.454	5.184
21	.294	.520	.154	.480	.056	.000	.217	.083	.130	.000	.300	.059	.059	2.352
22	.882	.429	.800	.250	.000	.522	.875	.545	.000	.500	.308	.273	.111	5.495
23	.875	.444	.625	.750	.500	.778	.333	.000	.789	.375	.000	.000	.091	5.560
24	.250	.778	.400	.643	.800	.375	.167	.500	.706	.111	.000	.000	.000	4.730
25	.857	.769	.500	.444	.067	.400	.882	.714	.200	.571	.000	.000	.125	5.529
26	.867	.750	.250	.000	.428	.714	.500	.333	.333	.500	.222	.333	.556	5.786
TOTALS	14.658	12.154	10.197	9.716	10.960	10.916	11.577	10.719	10.450	9.588	6.300	4.905	5.084	127.224

Table B-5

IDENTIFICATION ACCURACY SCORES: SUBJECT BY IMAGE

I... Code No.	1	2	3	4	5	6	Image Number 7	8	9	10	11	12	13	SUM
1	.455	.200	.417	.333	.400	.737	.556	1.000	.500	1.000	.250	.231	.615	6.694
2	.462	.083	.588	.333	.619	.400	.667	.000	.125	.500	.636	.556	.579	5.548
3	.357	.783	.200	.333	.450	.200	.385	.000	.000	.000	.500	.556	.312	4.076
4	.375	.355	.111	.176	.280	.385	.294	.000	.000	.077	.364	.600	.600	3.617
5	.461	.321	.625	.214	.364	.500	.200	.538	.000	.125	.444	.536	.684	5.012
6	.545	.444	.667	.444	.000	.304	.100	.125	.000	.000	1.000	.250	.647	4.526
7	.111	.100	.273	.222	.071	.500	.083	.182	.091	.000	.333	.640	.619	3.225
8	.545	.350	.500	.273	.267	.600	.600	.000	.000	.000	.500	.727	.684	5.046
9	.080	.242	.154	.077	.567	.207	.438	.357	.071	.000	.231	.278	.273	2.975
10	.353	.476	.462	.154	.571	.308	.428	.444	.000	.143	.556	.667	.842	5.404
11	.545	.174	.600	.267	.500	.200	.412	.333	.000	.125	.400	.304	.667	4.527
12	.400	.182	.385	.454	.364	.062	.364	.667	.000	.200	.286	.700	.650	4.714
13	.438	.143	.727	.625	.500	.778	.444	.857	.200	.000	.250	.800	.941	6.703
14	.600	.850	.500	.454	.478	.379	.462	1.000	.000	.250	.800	.727	.571	7.071
15	.583	.000	.000	.444	.571	.308	.636	.143	.400	.000	.833	.577	1.000	5.495
16	.000	.222	.111	.182	.167	.562	.667	.750	.000	.500	.714	.652	.800	5.327
17	.214	.310	.391	.286	.048	.475	.048	.000	.000	.100	.500	.543	.824	3.739
18	.571	.103	.500	.200	.333	.667	.235	.000	.000	.000	.500	.480	.522	4.111
19	.214	.462	.200	.222	.500	.091	.429	.444	.000	.000	.500	.783	.933	4.778
20	.438	.318	.583	.333	.474	.462	.308	.454	.111	.167	.400	.682	.454	5.184
21	.059	.059	.294	.083	.217	.520	.130	.056	.000	.000	.154	.480	.300	2.352
22	.250	.308	.273	.111	.882	.522	.545	.875	.000	.000	.500	.429	.800	5.495
23	.444	.625	.750	.000	.000	.789	.091	.875	.000	.333	.500	.778	.375	5.560
24	.375	.111	.778	.400	.643	.500	.000	.000	.000	.250	.167	.706	.800	4.730
25	.200	.067	.400	.571	.769	.125	.500	.444	.000	.000	.857	.714	.882	5.529
25	.333	.500	.333	.428	.714	.222	.500	.750	.250	.000	.333	.556	.867	5.786
TOTALS	9.408	7.788	10.822	7.619	10.749	10.803	9.522	10.294	1.748	3.770	12.508	14.952	17.241	127.224

Table B-6

ANALYSIS OF VARIANCE SUMMARY: IDENTIFICATION ACCURACY SCORES

Source of Variation	Sum of Squares	df	Mean Square	F	$F_{.95}$	$F_{.99}$
Between:						
Subjects	2.47175	25	.09887	3.20**	1.55	1.84
Within:						
Images	7.79496	12	.64958	21.00**	1.79	2.25
Conditions	3.64678	12	.30390	9.83**	1.79	2.25
Residual	8.90782	288	.03093	----	----	----
TOTAL	22.82131	337	------	----	----	----

**Means significantly different, $P \leq .01$.

Table B-7

IDENTIFICATION COMPLETENESS SCORES: SUBJECT BY TREATMENT CONDITION

Ind. Code No.	Treatment Condition Code 1-1-1	2-1-1	3-1-1	4-1-1	1-2-1	1-3-1	1-1-2	1-1-3	1-1-4	1-3-4	4-1-4	4-3-1	4-3-4	SUM
1	.500	.424	.500	.385	1.000	.316	.200	.417	.500	.857	.296	.077	.125	5.597
2	.875	.600	.303	.500	.769	.500	.667	.684	.500	.667	.000	.407	.038	6.510
3	.692	.500	.500	.121	.000	.154	.474	.000	.417	.333	.333	.000	.185	3.709
4	.556	.423	.500	.600	.303	.000	1.000	.154	.368	.417	.700	.500	.000	5.521
5	1.000	.481	.346	.500	.600	.424	.385	.000	1.000	.421	.250	.500	.500	6.407
6	.667	.143	.407	.154	.625	.600	.000	.212	.462	.000	.000	.083	.233	3.586
7	.533	.333	.286	.481	.077	.375	.424	.100	.500	.231	.000	.053	.083	3.476
8	.500	.533	.500	.000	.481	.269	.600	.625	.364	.000	.538	.000	.210	4.620
9	.895	.583	.333	.167	.714	.222	.375	.308	.200	.182	.500	.154	.000	4.633
10	1.000	.632	.500	.467	.667	.571	.385	.592	.625	.600	.121	.000	.462	6.622
11	.462	1.000	.526	.583	.233	.667	.370	.428	.154	.500	.600	.091	.000	5.614
12	.000	.385	1.000	.421	.333	.467	.857	.833	.481	.154	.250	.400	.030	5.611
13	.636	.500	.615	.000	.421	.333	.833	.533	.857	.592	.077	.125	.700	6.222
14	.654	1.000	.333	.500	.500	1.000	.833	.462	.579	.000	.533	.296	.300	6.990
15	.667	.625	.500	.556	1.000	.000	.583	.632	.143	.242	.700	.000	.000	5.648
16	.500	.444	.000	.077	.273	.625	.000	.857	1.000	.500	.077	.333	.053	4.739
17	.000	.346	.692	.667	.633	.518	.576	1.000	.500	.300	.053	.083	.000	5.368
18	.667	.667	.526	.333	.400	.115	.400	.625	.444	.462	.000	.000	.000	4.639
19	.600	.500	.571	.000	.154	.333	.300	.518	.462	.368	.000	.030	.125	3.961
20	.700	.500	1.000	.364	.474	.333	.538	.269	.667	.714	.250	.500	.370	6.679
21	.385	.394	.250	.400	.143	.000	.263	.333	.250	.000	.222	.100	.038	2.778
22	.789	.200	.444	.300	.000	.364	1.000	.500	.000	.375	.154	.231	.167	4.524
23	1.000	.400	.385	.462	.625	.467	1.000	.000	.454	.222	.000	.000	.083	5.098
24	1.000	.538	.667	.474	.444	.300	.125	.273	.400	.077	.000	.000	.000	4.298
25	.750	.526	.500	.571	.038	.154	.556	.500	.100	.667	.000	.000	.030	4.392
26	.481	.857	.500	.000	.500	.526	.346	.400	.231	.417	.121	.250	.167	4.796
TOTALS	16.509	13.534	12.684	9.083	11.407	9.633	13.090	11.255	11.658	9.298	5.775	4.213	3.899	132.038

Table B-8

IDENTIFICATION COMPLETENESS SCORES: SUBJECT BY IMAGE

Ind. Code No.	Image Number 1	2	3	4	5	6	7	8	9	10	11	12	13	SUM
1	.500	.077	.385	.500	.316	.424	.417	.857	.500	1.000	.125	.200	.296	5.597
2	.600	.038	.769	.667	.684	.303	.667	.000	.500	.500	.875	.500	.407	6.510
3	.500	.692	.154	.333	.474	.121	.417	.000	.000	.000	.500	.333	.185	3.709
4	.600	.423	.154	.500	.368	.303	.417	.000	.000	1.000	.500	.700	.556	5.521
5	.600	.346	.385	.500	.421	.424	.250	1.000	.000	1.000	.500	.500	.481	6.407
6	.600	.154	.462	.667	.000	.212	.083	.143	.000	.000	.625	.233	.407	3.586
7	.100	.077	.231	.333	.053	.424	.083	.286	.500	.000	.375	.533	.481	3.476
8	.600	.269	.538	.500	.210	.364	.500	.000	.000	.000	.625	.533	.481	4.620
9	.200	.308	.154	.167	.895	.182	.583	.714	.500	.000	.375	.333	.222	4.633
10	.600	.385	.462	.667	.632	.121	.500	.571	.000	1.000	.625	.467	.592	6.622
11	.600	.154	.462	.667	.526	.091	.583	.428	.000	1.000	.500	.233	.370	5.614
12	.400	.154	.385	.833	.421	.030	.333	.857	.000	1.000	.250	.467	.481	5.611
13	.700	.077	.615	.833	.421	.636	.333	.857	.500	.000	.125	.533	.592	6.222
14	.300	.654	.462	.833	.579	.333	.500	1.000	.000	1.000	.500	.533	.296	6.990
15	.700	.000	.000	.667	.632	.242	.583	.143	1.000	.000	.625	.500	.556	5.648
16	.000	.077	.077	.333	.053	.273	.500	.857	.000	1.000	.625	.500	.444	4.739
17	.300	.346	.692	.667	.053	.576	.083	.000	.000	1.000	.500	.633	.518	5.368
18	.400	.115	.462	.667	.526	.667	.333	.000	.000	.000	.625	.400	.444	4.639
19	.300	.462	.154	.333	.368	.030	.500	.571	.000	.000	.125	.600	.518	3.961
20	.700	.269	.538	.667	.474	.364	.333	.714	.500	1.000	.250	.500	.370	6.679
21	.100	.038	.385	.333	.263	.394	.250	.143	.000	.000	250	.400	.222	2.778
22	.300	.154	.231	.167	.789	.364	.500	1.000	.000	.000	.375	.200	.444	4.524
23	.400	.385	.462	.000	.000	.454	.083	1.000	.000	1.000	.625	.467	.222	5.098
24	.300	.077	.538	.667	.474	.273	.000	.000	.000	1.000	.125	.400	.444	4.298
25	.100	.038	.154	.667	.526	.030	.500	.571	.000	.000	.750	.500	.556	4.392
26	.400	.346	.231	.500	.526	.121	.417	.857	.500	.000	.250	.167	.481	4.796
TOTALS	10.900	6.115	9.542	13.668	10.684	7.756	9.748	12.569	4.500	12.500	11.625	11.365	11.066	132.038

Table B-9

ANALYSIS OF VARIANCE SUMMARY: IDENTIFICATION COMPLETENESS SCORES

Source of Variation	Sum of Squares	df	Mean Square	F	$F_{.95}$	$F_{.99}$
Between:						
Subjects	2.35077	25	.09403	2.16**	1.55	1.84
Within:						
Images	3.21364	12	.26780	6.15**	1.79	2.25
Conditions	6.44704	12	.53725	12.34**	1.79	2.25
Residual	12.53695	288	.04353	----	----	----
TOTAL	24.54840	337	------	----	----	----

**Means significantly different, $P \leq .01$.

Table B-10

ANALYSIS OF VARIANCE SUMMARY:

DETECTION ACCURACY SCORES FOR TARGET ANNOTATIONS

Source of Variation	Sum of Squares	df	Mean Square	F	$F_{.95}$	$F_{.99}$
Between:						
Subjects	2.517965	25	.100719	2.92**	1.55	1.84
Within:						
Images	1.044532	12	.087044	2.52**	1.79	2.25
Conditions	3.406700	12	.283892	8.24**	1.79	2.25
Residual	9.926800	288	.034468	----	----	----
TOTAL	16.895997	337	-------	----	----	----

**Means significantly different, $P \leq .01$.

Table B-11

ANALYSIS OF VARIANCE SUMMARY:

DETECTION ACCURACY SCORES FOR NON-TARGET ANNOTATIONS

Source of Variation	Sum of Squares	df	Mean Square	F	$F_{.95}$	$F_{.99}$
Between:						
Subjects	10.013154	25	.400526	9.86**	1.55	1.84
Within:						
Images	4.103743	12	.341979	8.42**	1.79	2.25
Conditions	1.361924	12	.113494	2.79**	1.79	2.25
Residual	11.698229	288	.040619	----	----	----
TOTAL	27.177050	337	-------	----	----	----

**Means significantly different, $P \leq .01$.

Table B-12

DUNCAN'S MULTIPLE RANGE TEST FOR DIFFERENCES AMONG

TREATMENT MEANS FOR DETECTION ACCURACY PERFORMANCE

FOR TARGET ANNOTATIONS ONLY

Treat-ment Code		4-1-4	4-3-1	4-3-4	4-1-1	1-3-4	1-3-1	1-1-3	1-2-1	1-1-4	2-1-1	1-1-2	1-1-1	3-1-1	Shortest Significant Ranges		
	Means	.658	.691	.699	.824	.859	.912	.915	.916	.924	.930	.931	.935	.938	α = .05		α = .01
4-1-4	.658	---			.166	.201	.254	.257	.258	.266	.272	.273	.277	.280	R_2	.101	.133
4-3-1	.691		---		.133	.168	.221	.224	.225	.233	.239	.240	.244	.247	R_3	.106	.138
4-3-4	.699			---	.125	.160	.213	.216	.217	.225	.231	.232	.236	.239	R_4	.110	.142
4-1-1	.824				---										R_5	.112	.145
1-3-4	.859					---									R_6	.115	.147
1-3-1	.912						---								R_7	.116	.149
1-1-3	.915							---							R_8	.118	.151
1-2-1	.916								---						R_9	.119	.152
1-1-4	.924									---					R_{10}	.120	.153
2-1-1	.930										---				R_{11}	.121	.154
1-1-2	.931											---			R_{12}	.122	.155
1-1-1	.935												---		R_{13}	.122	.156
3-1-1	.938													---			

Significant mean differences are listed above the diagonal (Difference = Column Mean - Row Mean).
Statistical significance level of differences is indicated by an underscore where P $\leq$ .01 and without an underscore where P $\leq$ .05.

Table B-13

DUNCAN'S MULTIPLE RANGE TEST FOR DIFFERENCES AMONG TREATMENT MEANS FOR DETECTION ACCURACY PERFORMANCE FOR NON-TARGET ANNOTATIONS ONLY

Treat-ment Code	Means	4-3-1	4-3-4	3-1-1	4-1-4	1-2-1	4-1-1	1-3-4	1-1-2	1-1-4	1-1-3	2-1-1	1-1-1	1-3-1	Shortest Significant Ranges	α = .05	α = .01
	Means	.452	.499	.500	.556	.569	.577	.582	.601	.616	.631	.633	.666	.667			
4-3-1	.452	---				.117	.125	.130	.149	.164	$\underline{.179}$	$\underline{.181}$	$\underline{.214}$	$\underline{.215}$	R_2	.110	.144
4-3-4	.499		---								.132	.134	$\underline{.167}$	$\underline{.168}$	R_3	.115	.150
3-1-1	.500			---							.131	.133	$\underline{.166}$	$\underline{.167}$	R_4	.119	.154
4-1-4	.556				---										R_5	.122	.157
1-2-1	.569					---									R_6	.124	.160
4-1-1	.577						---								R_7	.126	.162
1-3-4	.582							---							R_8	.128	.163
1-1-2	.601								---						R_9	.129	.165
1-1-4	.616									---					R_{10}	.130	.166
1-1-3	.631										---				R_{11}	.131	.167
2-1-1	.633											---			R_{12}	.132	.168
1-1-1	.666												---		R_{13}	.133	.169
1-3-1	.667													---			

Significant mean differences are listed above the diagonal (Difference = Column Mean - Row Mean).

Statistical significance level of differences is indicated by an underscore where P $\leq$ .01 and without an underscore where P $\leq$ .05.